Effects of syllable structure on intonation identification in Neapolitan Italian

Caterina Petrone, Mariapaola d'Imperio

Effects of syllable structure on intonation identification in Neapolitan Italian

Caterina Petrone[1, 2] and Mariapaola D'Imperio[1, 2,3]
[1] Laboratoire Parole et Langage, CNRS, [2] Aix-Marseille University, [3]Institut Universitaire de France

ABSTRACT

In Neapolitan Italian, nuclear rises are later in yes/no questions (L*+H) than in narrow focus statements (L+H*). Also, the H target is later in closed syllable items than in open syllable ones. In three identification tasks, we found that, when stimuli are ambiguous between questions and statements, listeners exploit the information on the precise alignment within the syllable to identify sentence type. This effect depends on durational constraints, i.e., the perceptual location of the H target is calculated relative to the actual duration of the vowel. Our results suggest that phonetic variability plays a role in shaping intonational categories and support models in which segmental and prosodic information are processed in a parallel fashion.

Keywords: intonation, tonal alignment, syllable structure, perception, Neapolitan Italian.

1. INTRODUCTION

In intonational studies, the search for primitives often coincided with the search of "invariant" f0 dimensions (such as slope or alignment). In particular, a hypothesis which has been widely discussed in intonational studies concerns the "segmental anchoring" of level tones, i.e., the idea that Low (L) and High (H) targets are systematically aligned with specific segmental boundaries or prosodic units. This stability has been taken as evidence that level tones (and not f0 movements) are the basic units of intonation representation [2, 3]. Despite these findings, other studies report that tonal alignment actually varies with a number of phonetic and phonological factors. In Neapolitan Italian, yes/no questions (Q) and narrow focus statements (S) have both a terminal falling contour and their distinction is manly cued by a small difference in the alignment the nuclear accent (around 40 ms). Namely, the nuclear LH rise is later in yes/no questions (L*+H) than in narrow focus statements (L+H*), and this phonetic difference is exploited for intonation identification [4]. Also, though vowel duration is not contrastive in Italian, stressed vowels are longer in open than in closed syllables, especially in penultimate position within the word [5]. [5] reported that the alignment of the H target of LH nuclear rises is later in closed syllables than in open syllables. As a consequence, the H target of L*+H and L+H* accents is differently located in lexical minimal pairs contrasting for consonantal length and syllable structure, such as *nono* (/ˈnono/) "ninth" and *nonno* (/ˈnɔnno/) "grandpa." In this minimal pair, the first word contains a singleton consonant and its stressed syllable is open (*no*-); the second word contains a geminate consonant and its stressed syllable is closed (*non*-). In fact, consonantal length is phonological in Italian. Also, the stressed vowel is shorter in closed than in open syllables. This interaction between intonation and syllable structure results in a four-way alignment pattern. In statements, the H target is realized at the vowel onset in open syllables; and slightly later in closed ones. In questions, it is realized around the vowel offset in open syllables; and in the following coda in closed ones [5].

A question related to the effect of syllable structure on tonal alignment concerns its nature, i.e., whether it is structural or if it depends on durational constraints. Evidence is quite mixed in the literature. In production, [8] found that, in Dutch, the prenuclear H aligns differently depending on the phonological length of the accented vowel. The difference was found even despite a lack of durational difference, thus supporting the idea that the H target seeks an alignment to the right edge of the syllable. However, when vowels have similar duration, alignment differences are very small, and speaker-specific differences are also found.

The current paper investigates the effects of the interaction between syllable structure and tonal alignment in Neapolitan Italian on the identification of the Q/S contrast. First, if listeners are capable of exploiting their knowledge on tonal alignment variability across syllable structure, scores for the Q/S contrast will differ when the accented word will contain an open vs. a closed stressed syllable. Moreover, we hypothesize that there will be no difference in intonation identification across syllable structure when the segmental duration is neutralized.

2. CORPUS AND METHODS

Three identification tasks were performed. Task 1 is a preliminary task. It helped in the creation of lexical stimuli whose durational values were employed as bases for the synthetic manipulations in the main tasks, i.e., Task 2 and Task 3. Durational values for items which were clearly identified in Task 1 as containing a closed (*nonno*) vs. open (*nono*) stressed

syllable were used as bases for alignment manipulations in Task 2; durational values for stimuli which were judged ambiguous in Task 1 as for the syllable structure (i.e., resulting in an ambiguous lexical identification between *nonno* and *nono*) were used as bases for alignment manipulations in Task 3. For each task, we created 2 continua of manipulation steps, one from a natural closed syllable (*nonno*) base and the other one from a natural open syllable (*nono*) base. In Task1, the natural items were produced by a Neapolitan speaker as citation forms. In Task 2 and 3, they were inserted in short carrier sentences (*é il TARGET*, "is it TARGET") and produced as yes/no questions (Q) or narrow focus statements (S) by the same speaker. The stressed syllable contained a nuclear rise (L*+H for Q; L+H* for S) followed by a terminal fall (L-L% in both sentence types). The stimuli were all resynthesized through PSOLA. Detailed duration and alignment manipulations are described below.

2.1. Stimuli

2.1.1. Task 1

The duration of the coda consonant (C) and that of the preceding vowel (V) within the first (stressed) syllable were both modified to investigate the impact of length cues to syllable structure on lexical identification. This type of manipulation is based on previous findings on consonantal gemination, suggesting that consonant duration is inversely correlated with the duration of the preceding vowel [9]. To avoid the presence of f0 related cues, the melodic contour was first stylized so as to obtain an evenly falling contour ranging from 170 to 130 Hz, with values averaged over the two lexical bases. Once the f0 contour was stylized, the duration of the target word containing the open syllable base (*nono)* was manipulated to create a continuum of 9 steps (D1-D9). That is, we simultaneously modified the stressed vowel and the following coda consonant. We used steps of 10 ms to progressively decrease vowel duration and of 15 ms to progressively increase the following consonant duration. We then performed the same manipulations on the closed syllable base (*nonno)* and created a 9 steps continuum in the same way and with same step sizes as for the open syllable base. That is, we increased the duration of the stressed vowel (V) and at the same time decreased that of the coda consonant (C) from step D9 to D1. By doing so, we obtained values similar to the original productions at fixed steps of the continuum (D2 and D9). At D2, V was 240 ms long and C was 85 ms long. At D9 of the continuum, V was 180 ms long while C was 175 ms long (see Table 1).

Table 1: Steps D1-D9 for the duration manipulation (ms) on the stressed vowel (V) and on the following consonant (C).

	D1	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**
V	250	240	230	220	210	200	190
C	70	85	100	115	130	145	160

2.1.2. Task 2

In this task, we investigated the impact of tonal alignment on the Q/S contrast identification when the accented syllable has a clear durational difference across the closed vs. open syllable items. As base for the manipulation, we modified the duration of the two base stimuli by using the same durational values as those of stimuli judged unambiguous by participants in Task 1, i.e. at step D2 for the open (*nono*) and at step D8 for the closed syllable (*nonno*) item. Each stimulus served as the basis for a tonal alignment manipulation in 14 steps of 15 ms for both Q and S bases (A1-A14). Prior to the manipulation, f0 height differences were neutralized, by imposing melodic values that were intermediate between the two sentences types. The f0 contour was then stylized by fitting straight lines across 5 points: 1 point at the start and 1 at the end of the utterance; 3 points for each tonal target in the nuclear rise-fall configuration, L1 (the start of the nuclear rise), H (the nuclear peak) and L2 (the phrase accent). The three targets (L1, H and L2) were simultaneously shifted left- or rightwards relative to vowel offset, which allowed to obtain a continuum from an early peak alignment typical of statements to a late peak alignment typical of questions (Table 2).

Table 2: Steps A1-A14 for the alignment manipulation (ms) of the nuclear configuration.

	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**
L1	-305	-290	-275	-260	-245	-230	-215
H	-135	-102	-105	-90	-75	-60	-45
L2	35	50	65	80	95	110	125

	A8	**A9**	**A10**	**A11**	**A12**	**A13**	**A14**
L1	-200	-185	-170	-155	-140	-125	-110
H	-30	-15	0	15	30	45	60
L2	140	155	170	185	200	215	230

2.1.3. Task 3

Task 3 aimed at investigating the perception of the Q/S contrast when the duration of the accented syllable is perceptually ambiguous. Specifically,

from Task 1, we obtained durational values of perceptually ambiguous stimuli, corresponding to stimuli at D5 (for) and D6 (for). Both tone height neutralization and subsequent linear stylization followed the same procedure as in Task 2. Subsequently, the tonal alignment of the three targets (L1-H-L2) was manipulated by using the alignment values reported in Table 2.

2.2. Participants

The three tasks were run with three separate groups of listeners, i.e., 7, 20 and 22 listeners respectively for Task 1, 2 and 3. They were native speakers of Neapolitan Italian, aged 20-30 y.o. and were all right-handed.

2.3. Procedure

In the three tasks, listeners were asked to perform a two-alternative forced choice task by pressing one of two corresponding buttons on a button box. In Task 1, listeners had to perform a lexical identification task (*nonno* or *nono*), while in Tasks 2 and 3 they performed an intonation contour identification task (*question* or *statement*). Each task was preceded by a short practice session and it lasted around 20 minutes. The setup was the same for all tasks. The auditory stimuli were played from a laptop through the software PERCEVAL [1]. Each stimulus was heard 7 times in 7 randomized blocks for Task 1. Stimuli were heard 2 times in Tasks 2 and 3. The statistical analysis included a series of logit models with mixed effects separately for each task, in which the bases and the steps of manipulation were the fixed factors and listeners were the random intercept ($p < .01$).

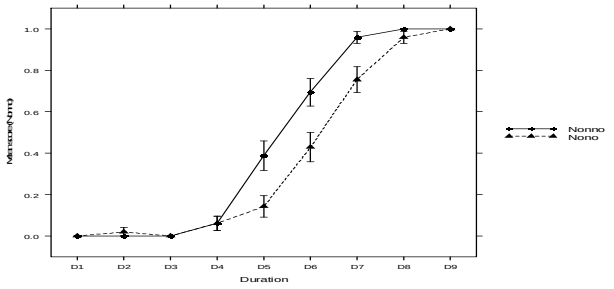
To sum up, we obtained 882 obs. for Task 1 (2 lexical bases * 9 steps * 7 times * 7 listeners), 1120 obs. for Task 2 (2 intonation bases*14 steps*2 times*20 listeners) and 1232 obs. for Task 3 (2 intonation bases *14 steps*2 times*22 listeners).

3. RESULTS

3.1.1. Task 1

Fig. 1 shows the mean score for the identification of stimuli from the closed syllable base (*nonno*, dots) and the open syllable base (*nono*, triangles). Responses varied with duration manipulations. At early locations within the continuum (D1-D4), responses for the open syllable item were dominant, while at the opposite end of the continuum (D9-D10), responses for the closed syllable item were dominant. The figure also suggests a base stimulus effect, since the response function is shallower stimuli created from the open (*nono*) than from the closed syllable (*nonno*) base. The statistical analysis confirmed both an effect of step manipulation [z = 13.71] and of lexical base [z = -4.42].

Figure 1: Mean *nonno* score for the two continua created from the *nonno* and *nono* bases.



3.1.2. Task 2

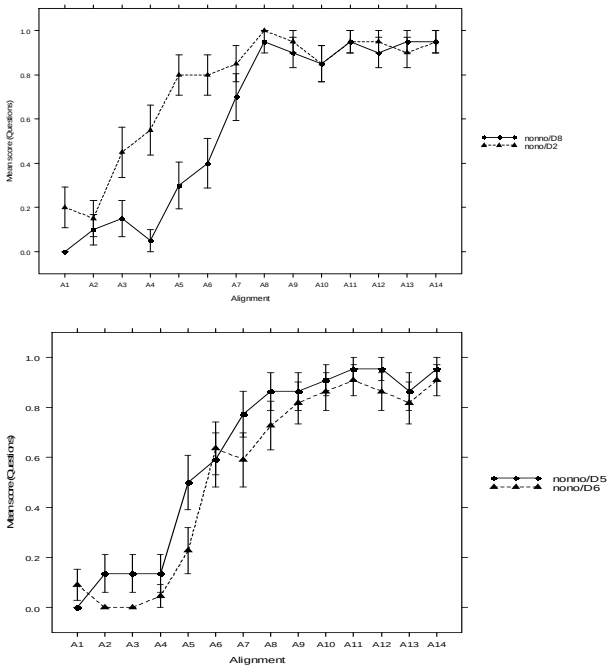
In Fig. 2a-2b, results for Q/S identification are presented for all listeners. For lack of space, data are shown only for the question-base continuum (results for the statement-base continuum were similar). In Fig. 2a, mean score for questions (y axis) is plotted against alignment steps (x axis), separately for stimuli created from the open syllable base *nono*/D2 (triangles) and from closed syllable base *nonno*/D8 (dots) lexical bases. In line with findings in [3,4], question identification score increased when the entire rise-fall (L1, H and L2) was shifted rightwards (e.g., at A9-14), while it decreased when the targets were shifted leftwards (e.g., at A1-A2). While at the extrema of the continua, the two curves overlap, in the middle of the continua (A3-A8) the response function for the Q/S identification are modulated by syllable structure. For instance, at A3 the "question" score was already close to chance level for the open syllable base *nono*/D2, for both question and statement base continua. However, the identification score at the same alignment step was much lower for the closed syllable base *nonno*/D8. The effect of manipulation steps [z = 10.6] and syllable structure [z = 4.52] were both significant. None of the interactions nor sentence base type were significant.

3.1.3. Task 3

Fig. 2b shows mean score for questions (y axis) against alignment steps (x axis), separately for stimuli created from the open syllable base *nono*/D6 (triangles) and from the closed syllable base *nonno*/D5 (dots). As for Task 2, the identification score was a function of the tonal alignment manipulation. Moreover, the score was very similar for both lexical bases (*nonno*/D5 and *nono*/D6). The logit regression confirmed a significant effect of manipulation steps [z = 11.14, p <.01], while neither

the lexical nor the intonational bases were significant. The interactions were not significant.

Figure 2: Mean question score for the continua created from the question bases and containing unambiguous (2a, upper) and ambiguous (2b, lower) accented syllable structure.



4. DISCUSSION

Task 1 confirms previous results on Italian gemination, i.e., both consonant and preceding vowel duration are strong cues to distinguish lexical minimal pairs contrasting in consonantal length syllable structure [6]. The effect of lexical base suggests the existence of additional phonetic cues other than duration for consonantal gemination.

Results for Task 2 show that syllable structure affects intonation identification. The "question" identification score was globally higher when the accented syllable was open than when it was closed. Syllable structure effects were stronger at the middle of the alignment continua than at their extrema values. This means that Neapolitan listeners are able to exploit information about alignment variability for Q/S identification when intonation identification is ambiguous. When the f0 nuclear peak was located in the very early portion of the syllable (i.e., at one extremum of the alignment continuum), listeners interpreted the stimuli as statements since an early alignment only characterizes statements in Neapolitan. When the f0 peak was at the end of the accented vowel or in the following consonant (at the other extremum of the alignment continuum), listeners interpreted the stimuli as questions, since such a late alignment pattern is a cue for question interpretation. In the region of ambiguity (i.e., in the middle of the continua) at steps A3-A8, the f0 peak was located in the middle portion of the stressed vowel, so that intonation identification was more difficult. Hence, listeners disambiguated the Q/S identification by computing the exact alignment of the f0 peak relative to syllable offset. The response function is steeper for the open syllable base than for the closed syllable base. This is congruent with acoustic detail, since in questions the H target is placed earlier in open than in closed syllable items.

Task 3 revealed that Neapolitans are not able to exploit syllable structure information when segmental duration is ambiguous between the closed and the open syllable items. This means that the effects found in Task 2 can be explained in terms of durational constraints, i.e., peak alignment is computed relative to the actual duration of the stressed vowel. According to this account, Neapolitan listeners would expect the f0 nuclear rise to employ a fixed amount of time in order to be realized: when the vowel is phonetically short (as in the closed syllable items), the f0 rise cannot be completed within the vowel boundaries, so that the H target is expected to be located within the following coda consonant; when the vowel is phonetically long (as in the open syllable items), the rise can be achieved in the vowel, so that listeners expect the H target to be realized within the vowel boundaries. However, when precise duration information is not available, Neapolitans are not able to compute perceptual alignment differences in closed vs. open syllable items.

Our results are interpretable within models which take into account phonetic detail in shaping linguistic categories [7]. If linguistic categories are "centers of density" [10], phonetic detail would play a minor role within the center of the category, while it would influence boundary location at the periphery of the category. Hence, for the purpose of the Q/S contrast, segment duration might have no effect in the center of the intonational category (since tonal alignment is the main cue for the L+H* vs. L*+H contrast), while it would be exploited in the rarified area of the category (when alignment is ambiguous between the two categories).

5. CONCLUSION

Our results suggest that segmental duration can be exploited for identifying intonational categories. This supports the idea that phonological categories are characterized by a rich internal structure. Finally, our data run counter to models that do not predict any interaction between the segmental and the suprasegmental level of speech.

7. REFERENCES

[1] André, C., Ghio, A., Cavé, C., Bernard, T. (2005). PERCEVAL (version 3.0.3). Retrieved from http: //www.lpl.univ-aix.fr/~lpl/dev/Perceval

[2] Arvaniti, A., Ladd, D. R., Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *J. Phon* 26, 3-25.

[3] D'Imperio, M. (2011). Prosodic representations. In A. Cohn, C. Fougeron & M. Huffman (Eds.), *Handbook of Laboratory Phonology*. Oxford: Oxford University Press.

[4] D'Imperio, Mariapaola, and House, David (1997). Perception of questions and statements in Neapolitan Italian. In *Proc. Eurospeech'97*, vol. 1. 251-254.

[5] D'Imperio, M., Petrone, C., and Nguyen, N. (2007). Effects of tonal alignment on lexical identification in Italian. In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes*. The Hague: Mouton de Gruyter, 79-106.

[6] Esposito, A., and Di Benedetto, M. G. (1999). Acoustic and perceptual analysis of Italian stop consonants. *JASA* 106, 2051-2062.

[7] Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *J. Phon* 31, 373-405.

[8] Ladd, D. R., Mennen, I., and Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *JASA* 107, 2685-2696.

[9] Mattei, M., and Di Benedetto, M. G. (2000). Acoustic analysis of singleton and geminate nasals in Italian. *Europ. Student J. of Language and Speech*. Retrieved from http: //www.essex.ac.uk/.

[10]Post, B., D'Imperio, M., and Gussenhoven, C. (2007). Fine phonetic detail and intonational meaning. *Proc. of the XVI ICPhS*, 191-196.

.